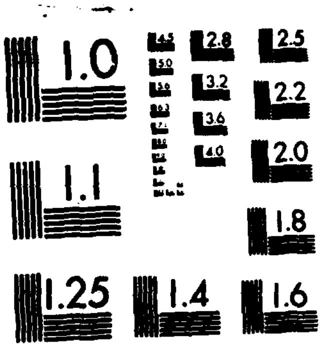
AD-A129 082 AN EMPIRICAL INVESTIGATION OF SEVERAL TESTS FOR THE MEAN OF A FIRST-ORDER..(U) DESMATICS INC STATE COLLEGE PA K C BURNS ET AL. MAY 83 TR-112-14 N00014-79-C-0128 1/1

UNCLASSIFIED F/G 12/1 NL

END
DATE
FILMED
7 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DESMATICS, INC.

P. O. Box 618
State College, Pa. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

AN EMPIRICAL INVESTIGATION OF SEVERAL TESTS FOR THE MEAN OF A FIRST-ORDER AUTOREGRESSIVE PROCESS

by
Kevin C. Burns
and
Dennis E. Smith

TECHNICAL REPORT NO. 112-14

May 1983

This study was supported by the Office of Naval Research under Contract No. N00014-79-C-0128, Task No. NR 207-037

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	STATISTICAL METHODS AND RESULTS	3
III.	SUMMARY	9
IV.	REFERENCES	10
	TABLES AND FIGURES	11

I. INTRODUCTION

The study of the effects of unusual environments on individuals often entails the analysis of repeated measurements taken on a single subject. Unfortunately, except under very restrictive sets of assumptions, no valid statistical techniques have been developed for such an analysis. In particular, for the first-order autoregressive process (AR(1)), no inference procedure is currently available which enables the analyst to control the probability of making an invalid conclusion. This problem is particularly acute when only a relatively small number of observations are available for the analysis. Four test statistics are considered for testing hypotheses about the mean of this autoregressive process,

The purpose of this investigation is to evaluate some of the procedures which have been suggested for this situation. Of particular interest is the difference between the nominal error rate chosen by the experimenter and the actual error rate given by the procedure. It is also desirable to evaluate how this difference is affected by the number of observations used in the analysis.

In a previous Desmatics technical report [1], Burns and Smith discussed the problem of testing hypotheses about the mean of an AR(1). Among the test statistics considered in that investigation were a modified t statistic, originally proposed by Higgins [2], and a more standard technique which involves transforming the observed data and treating the transformed observations as an independent sample. These two testing procedures are investigated more extensively in this report. In addition, both of these procedures require an estimate of the autocorrelation, and two such estimates are considered here. The first is the standard estimate, as given in any elementary

statistics text, while the second estimator includes a correction for bias. Thus, four procedures in all are considered in this investigation.

It should be noted here that the standard t statistic, which gives a valid test procedure only when the observations are independent, is not included in this investigation. The principal reason for this omission is that this procedure performs substantially worse, for any size sample, than any of the procedures which are considered. (This was shown clearly in the previous technical report.) Furthermore, since both of the estimators used here for the autocorrelation are consistent, the four procedures being considered are at least asymptotically valid, while the standard t statistic gives very poor results even in the asymptotic case.

II. STATISTICAL METHODS AND RESULTS

The problem considered here is that of testing hypotheses about the mean of a first-order autoregressive process. The model may be specified as follows:

$$Y_1 - \mu = \varepsilon_1 ,$$

$$Y_t - \mu = \rho(Y_{t-1} - \mu) + \varepsilon_t, \ t=2,3,\ldots,n \text{ where}$$

$$\varepsilon_1 \sim N(0,\sigma_\varepsilon^2/(1-\rho^2)),$$

$$\varepsilon_t \sim N(0,\sigma_\varepsilon^2) \text{ for } t \geq 2 \text{ and}$$

$$\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n \text{ are independent.}$$

Thus, the correlation between any two observations Y_i and Y_j is $\rho^{|i-j|}$. Also, $Y_t \sim N(\mu,\sigma^2)$ for all t where $\sigma^2 = \sigma_\varepsilon^2/(1-\rho^2)$.

As in the earlier technical report, attention is restricted to testing the hypothesis H_0: $\mu = 0$ vs. H_A: $\mu > 0$. As mentioned in that report, procedures used for testing this hypothesis may easily be extended to more complicated situations, such as tests concerning an intervention effect.

If $\rho = 0$, the problem given above reduces to that of testing whether $\mu = 0$ in a normal distribution. The observations are independent and the appropriate test statistic is $T = \sqrt{n}\,\bar{Y}/s$, which follows Student's t distribution with n-1 d.f. When $\rho > 0$, use of this statistic leads to a seriously inflated type I error rate. (See, for example, [1] or [2].) This inflated error rate is primarily a result of the fact that s^2 underestimates the variance of $\bar{Y}$, which is approximately $(\sigma^2/n)\left(\frac{1+\rho}{1-\rho}\right)$.

This approximation led Higgins to suggest the modified statistic $TC = \left(\frac{1-\hat{\rho}}{1+\hat{\rho}}\right)^{\frac{1}{2}} T$, where $\hat{\rho}$ is an appropriate estimate of the autocorrelation. Another alternative test statistic may be obtained by considering the trans-

formation $Z_t = Y_t - \rho Y_{t-1}$. When ρ is known, this transformation yields a set of n-1 independent observations from a normal distribution with mean $\nu = (1-\rho)\mu$. Since $\nu = 0$ when $\mu = 0$ and $\nu > 0$ when $\mu > 0$, the hypothesis H_0: $\nu = 0$ vs. H_A: $\nu > 0$ is equivalent to the original hypothesis. The appropriate test statistic is $TR = \sqrt{n-1}\,\overline{Z}/s_Z$. Of course, in practical situations, ρ is not known. Therefore, some estimate of the autocorrelation, $\hat{\rho}$, must be used for the transformation. Obviously, this procedure will only be as good as the estimate used.

The standard estimate of the first-order serial correlation is:

$$\hat{\rho}_1 = \{ \sum_{i=1}^{n-1} [(Y_i - \overline{Y})(Y_{i+1} - \overline{Y})] / \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \}.$$

Unfortunately, this estimate of the autocorrelation is biased, especially when the number of observations is small. A less biased estimate, which is derived in [3], is $\hat{\rho}_2 = [(n-1)\hat{\rho}_1 + 1]/(n-4)$. Since the range of $\hat{\rho}_1$ is [-1,1], it is possible for $\hat{\rho}_2$ to have inadmissable values (values outside of [-1,1]), particularly when n is small. In that case, $\hat{\rho}_2 = 1$ (or -1) is used.

As mentioned earlier, four different test statistics are included in this investigation. Two of these statistics, denoted TC1 and TC2, are obtained by using $\hat{\rho}_1$ and $\hat{\rho}_2$, respectively, as the estimate of the autocorrelation when calculating TC. (TC1 is the test statistic studied by Higgins.) The other two statistics, denoted TR1 and TR2, are calculated by using $\hat{\rho}_1$ and $\hat{\rho}_2$ to transform the data and proceeding as described for TR above.

Five different autocorrelations (ρ = .5, .7, .8, .85, .9) and five sample sizes (n = 10, 20, 30, 50, 100) were used in this investigation. For each value of (ρ, n), 1000 samples were simulated and the four test statistics calculated for each sample. From the 1000 simulations, the empirical distribution function was found for each test statistic. Finally, using $t_\alpha(n-1)$

as the critical value, the empirical significance level was found using each of three different nominal significance levels ($\alpha = .05, .025, .01$). The values obtained in this way are given in Tables 1 through 5. (The predicted values, which are also given in the tables, will be discussed later.)

It should be noted here that the observed empirical significance levels presented in Tables 1 through 5 are only estimates of the significance levels which are obtained when using the procedures being discussed in this report. The variability of these estimates may be calculated by considering the method by which they were obtained. For each test statistic, the empirical significance levels were calculated by counting the number of times, out of 1000 simulations, the statistic exceeded a specified critical value. This quantity is a random variable having a binomial distribution. Therefore, if p is the true probability of exceeding the critical value, the standard deviation of this random variable is $[p(1-p)/1000]^{\frac{1}{2}}$. If $p = .05$, for example, the standard deviation is .0069 and a 95% confidence interval for the empirical significance level is (.036, .064). (The normal approximation to the binomial distribution is used here to compute the confidence interval. This is nearly exact for n=1000.)

The actual significance levels for each test statistic are expected to be monotone decreasing functions of sample size, since the precision of the estimate of ρ increases as the sample size increases. The fact that the values given in the table do not always follow this pattern is attributable to the statistical variation described above. For $\rho = .5$, in particular, the observed values tend to behave erratically. For larger values of ρ, the actual change in the significance level as a function of sample size is large enough to overwhelm any small fluctuations due to statistical variability.

Since the observed empirical significance levels behave somewhat erratically as a function of sample size, it is difficult to determine how best to interpolate between the values in the tables, or how a given value should perhaps be adjusted after consideration of the values nearest it. It was decided that the best way to accomplish both purposes was to fit a function to each set of five observed significance levels. (The five values are for the five sample sizes considered with both ρ and α fixed.) Since the values in Tables 1 through 5 appear to decrease approximately exponentially as functions of sample size, functions of that type were first considered. It was finally decided, however, that in order to achieve increased flexibility, gamma functions should be fit to the data. These are functions of the form:

$$\Gamma(n) = \beta_0 n^{\beta_1} e^{\beta_2 n}.$$

These are monotone decreasing functions as long as $\beta_1 < 0$ and $\beta_2 < 0$. In order for the functions to be asymptotically equal to the nominal significance level, functions of the form $\alpha + \Gamma(n)$ were actually fit to the data.

Least squares regression, applied to the log transformation of the data, was used to fit the functions given above. That is, the functions actually fit were of the form:

$$\ln\ (y - \alpha) = \ln\ \beta_0 + \beta_1\ \ln\ n + \beta_2 n,$$

where y is the observed significance level. Unfortunately, many (22 of 60) of the functions fit in this way did not satisfy the restriction that $\beta_2 < 0$. In those cases, the functions:

$$\ln\ (y - \alpha) = \ln\ \beta_0 + \beta_1 \ln n$$

were used. In all of the cases where they were used, the simple functions gave almost as good a fit as did the full gamma functions. (Simple exponential functions were also considered, but did not fit the data well.)

The fitted functions have been plotted and are presented in Figures 1 through 15. The predicted values for each function have also been calculated for each of the sample sizes used in the regressions. These predicted values are listed in Tables 1 through 5 so that they may be compared to the observed values.

Comparison of the observed and predicted values in the tables show that in most cases the functions fit the data remarkably well. The exception to this occurs when $\rho = .5$, particularly for n=10. However, as mentioned earlier, the observed values exhibit rather erratic behavior when $\rho = .5$ and the functions cannot be expected to fit well in this situation. Another fact which should be noted is that the predicted and observed values are closest for large n. This is to be expected since the fitting was done on the log scale, which gives added weight to small values.

From Figures 1 through 15, it is clear that the two test statistics using $\hat{\beta}_2$ perform substantially better, for all values of ρ, than the test statistics which use $\hat{\beta}_1$. Furthermore, there is little difference between TR and TC, using either estimate of ρ, although TC1 generally does slightly better than TR1 and TR2 generally does slightly better than TC2. As could be expected, all of the test statistics perform better for moderate autocorrelations than they do when the autocorrelation is very high. For TR2, for example, with nominal $\alpha = .05$ and $\rho = .5$, a sample size of about 12 is needed to obtain an estimated significance level of .075. When $\rho = .9$, a sample size of 100 gives the same estimated level of significance.

As an example of how Figures 1 through 15 might be used, suppose that 20 repeated measurements are taken on an individual and that those measurements are assumed to follow a first-order autoregressive process. From the estimated autocorrelation, in conjunction with any prior information, the

experimenter decides that ρ is in the interval (.7, .9) . Now suppose TR2 is used to test whether the mean of the process is zero with nominal $\alpha = .01$. From Figures 6 and 15 one can obtain rough bounds on the actual significance level. In this case, the bounds are (.040, .087) .

III. SUMMARY

Four test statistics have been considered as candidates for testing H_0: $\mu = 0$ vs. H_A: $\mu > 0$ when the observations are taken from an AR(1) with autocorrelation ρ. A set of 1000 samples was generated for each of five different sample sizes (n=10, 20, 30, 50, 100) and five different autocorrelations (ρ = .5, .7, .8, .85, .9). From the 1000 simulations, the empirical distribution functions were calculated for each test statistic. Finally, using $t_\alpha(n-1)$ as the critical value corresponding to a specified nominal significance level, the empirical significance levels were found and tabulated.

Since these empirical significance levels were found to fluctuate erratically due to statistical variation, smoothing functions were fit to the five values for each combination of ρ, α, and test statistic. These functions are also an aid in interpolation between sample sizes. The functions are presented graphically and an example given as to how they might be used in practice.

VI. REFERENCES

[1] Burns, K. C. and Smith, D. E., "The Effect of Environmental Change in Single-Subject Experiments", Technical Report No. 112-12, Desmatics, Inc., 1983.

[2] Higgins, J. J., "A Robust Model for Estimating and Testing for Means in Single Subject Experiments," Human Factors, Vol. 20, pp. 717-724, 1978.

[3] Kendall, M. G. and Stuart, A., The Advanced Theory of Statistics, Vol.3, Hafner Publishing Company, New York, 1968.

Sample Size	TC1 Observed	TC1 Predicted	TC2 Observed	TC2 Predicted	TR1 Observed	TR1 Predicted	TR2 Observed	TR2 Predicted
				$\alpha = .05$				
10	.140	.125	.095	.079	.154	.140	.095	.079
20	.086	.096	.059	.070	.099	.103	.064	.068
30	.089	.084	.071	.066	.089	.089	.067	.064
50	.068	.074	.062	.062	.069	.076	.053	.060
100	.068	.065	.060	.059	.070	.066	.064	.056
				$\alpha = .025$				
10	.102	.088	.077	.063	.121	.103	.073	.057
20	.051	.060	.040	.044	.059	.067	.040	.040
30	.058	.049	.044	.037	.058	.054	.038	.035
50	.034	.041	.027	.032	.037	.043	.026	.031
100	.037	.034	.033	.029	.038	.035	.034	.028
				$\alpha = .01$				
10	.075	.062	.052	.047	.080	.073	.050	.046
20	.032	.038	.026	.029	.037	.040	.028	.029
30	.032	.029	.026	.023	.033	.030	.023	.023
50	.018	.022	.015	.018	.018	.022	.016	.018
100	.019	.017	.015	.014	.017	.016	.015	.014

Table 1: Comparison of Observed Empirical Significance Levels With Those Predicted by Smoothing Function for Four Test Statistics: $\rho = .5$.

Sample Size	TC1 Observed	TC1 Predicted	TC2 Observed	TC2 Predicted	TR1 Observed	TR1 Predicted	TR2 Observed	TR2 Predicted
				$\alpha = .05$				
10	.174	.173	.108	.107	.184	.183	.099	.098
20	.121	.128	.088	.091	.135	.144	.085	.086
30	.120	.109	.085	.083	.138	.123	.077	.079
50	.087	.091	.075	.075	.095	.100	.074	.072
100	.075	.075	.065	.065	.075	.075	.063	.063
				$\alpha = .025$				
10	.127	.127	.081	.082	.143	.142	.078	.075
20	.094	.095	.066	.063	.105	.108	.060	.059
30	.081	.080	.051	.054	.092	.091	.048	.053
50	.063	.063	.046	.045	.072	.072	.047	.046
100	.045	.045	.035	.035	.050	.050	.040	.039
				$\alpha = .01$				
10	.099	.099	.067	.068	.105	.106	.064	.065
20	.063	.061	.041	.037	.071	.069	.043	.040
30	.044	.047	.026	.027	.051	.053	.028	.030
50	.036	.034	.019	.020	.039	.038	.022	.021
100	.023	.023	.015	.015	.024	.024	.014	.014

Table 2: Comparison of Observed Empirical Significance Levels With Those Predicted By Smoothing Function for Four Test Statistics: $\rho = .7$.

Sample Size	TC1 Observed	TC1 Predicted	TC2 Observed	TC2 Predicted	TR1 Observed	TR1 Predicted	TR2 Observed	TR2 Predicted
				$\alpha = .05$				
10	.203	.200	.139	.132	.235	.237	.146	.135
20	.147	.147	.090	.102	.169	.170	.089	.103
30	.131	.125	.107	.090	.162	.143	.105	.090
50	.095	.105	.072	.079	.102	.117	.071	.078
100	.089	.085	.070	.068	.098	.094	.070	.067
				$\alpha = .025$				
10	.163	.165	.106	.100	.189	.183	.107	.103
20	.111	.116	.066	.079	.123	.128	.066	.076
30	.110	.096	.082	.069	.112	.105	.077	.064
50	.069	.076	.055	.057	.076	.084	.050	.054
100	.058	.057	.043	.043	.066	.063	.044	.044
				$\alpha = .01$				
10	.125	.120	.079	.075	.150	.146	.081	.077
20	.075	.086	.049	.057	.089	.097	.047	.054
30	.079	.070	.052	.047	.081	.076	.048	.043
50	.050	.051	.038	.038	.057	.057	.034	.034
100	.031	.031	.026	.026	.038	.038	.023	.023

Table 3: Comparison of Observed Empirical Significance Levels With Those Predicted by Smoothing Function for Four Test Statistics: $\rho = .8$.

Sample Size	TC1 Observed	TC1 Predicted	TC2 Observed	TC2 Predicted	TR1 Observed	TR1 Predicted	TR2 Observed	TR2 Predicted
				$\alpha = .05$				
10	.273	.273	.197	.183	.311	.316	.177	.177
20	.192	.192	.125	.131	.229	.228	.118	.119
30	.158	.158	.104	.111	.197	.186	.100	.098
50	.126	.126	.095	.092	.136	.143	.080	.081
100	.094	.094	.077	.076	.099	.098	.067	.067
				$\alpha = .025$				
10	.237	.231	.155	.154	.257	.253	.141	.133
20	.144	.152	.099	.103	.173	.183	.081	.090
30	.121	.120	.085	.082	.156	.148	.072	.072
50	.092	.090	.061	.062	.108	.109	.058	.055
100	.061	.061	.043	.043	.065	.065	.038	.038
				$\alpha = .01$				
10	.195	.193	.117	.118	.213	.214	.113	.117
20	.115	.120	.074	.076	.138	.139	.073	.067
30	.093	.089	.062	.058	.109	.105	.046	.049
50	.059	.060	.038	.040	.070	.072	.033	.033
100	.032	.032	.024	.024	.040	.040	.021	.021

Table 4: Comparison of Observed Empirical Significance Levels With Those Predicted by Smoothing Function for Four Test Statistics: $\rho = .85$.

Sample Size	TC1 Observed	TC1 Predicted	TC2 Observed	TC2 Predicted	TR1 Observed	TR1 Predicted	TR2 Observed	TR2 Predicted
				$\alpha = .05$				
10	.269	.271	.192	.195	.300	.300	.175	.179
20	.232	.229	.156	.152	.262	.260	.143	.141
30	.202	.200	.130	.131	.224	.229	.131	.122
50	.158	.161	.108	.109	.185	.182	.094	.100
100	.106	.106	.085	.085	.114	.114	.076	.075
				$\alpha = .025$				
10	.235	.242	.143	.146	.273	.276	.138	.139
20	.197	.188	.125	.120	.220	.214	.115	.111
30	.164	.158	.106	.105	.178	.180	.091	.095
50	.113	.121	.082	.085	.138	.139	.075	.074
100	.075	.074	.058	.058	.088	.088	.048	.048
				$\alpha = .01$				
10	.208	.209	.120	.123	.237	.240	.105	.106
20	.154	.152	.105	.101	.180	.177	.088	.087
30	.124	.123	.088	.086	.145	.143	.076	.074
50	.089	.090	.061	.064	.101	.104	.055	.057
100	.052	.052	.036	.036	.058	.058	.033	.033

Table 5: Comparison of Observed Empirical Significance Levels With Those Predicted by Smoothing Function for Four Test Statistics: $\rho = .9$.

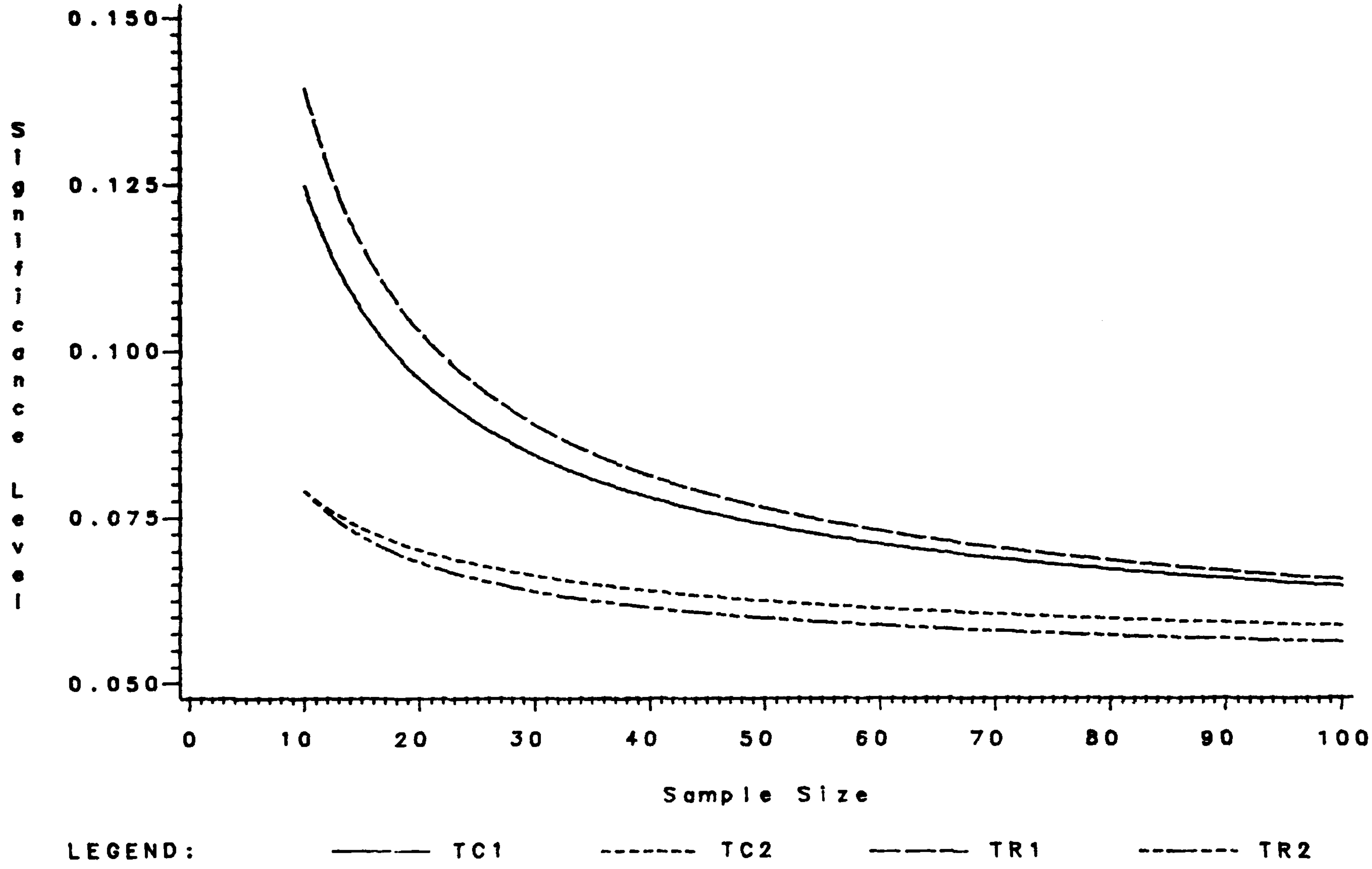


Figure 1: Empirical Significance Level as a Function of Sample Size for ρ=.5 and Nominal α=.05

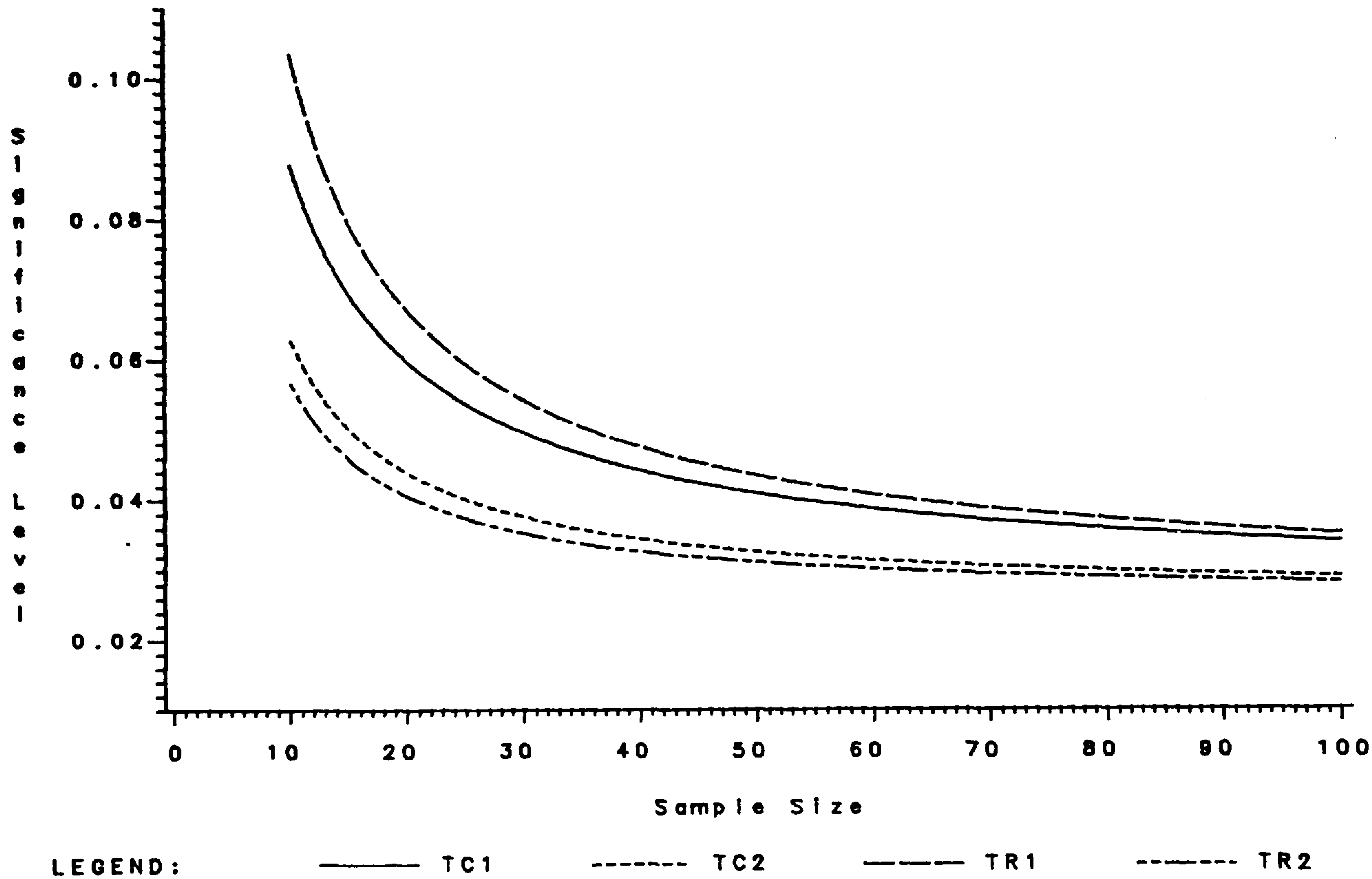


Figure 2: Empirical Significance Level as a Function of Sample Size for ρ=.5 and Nominal α=.025

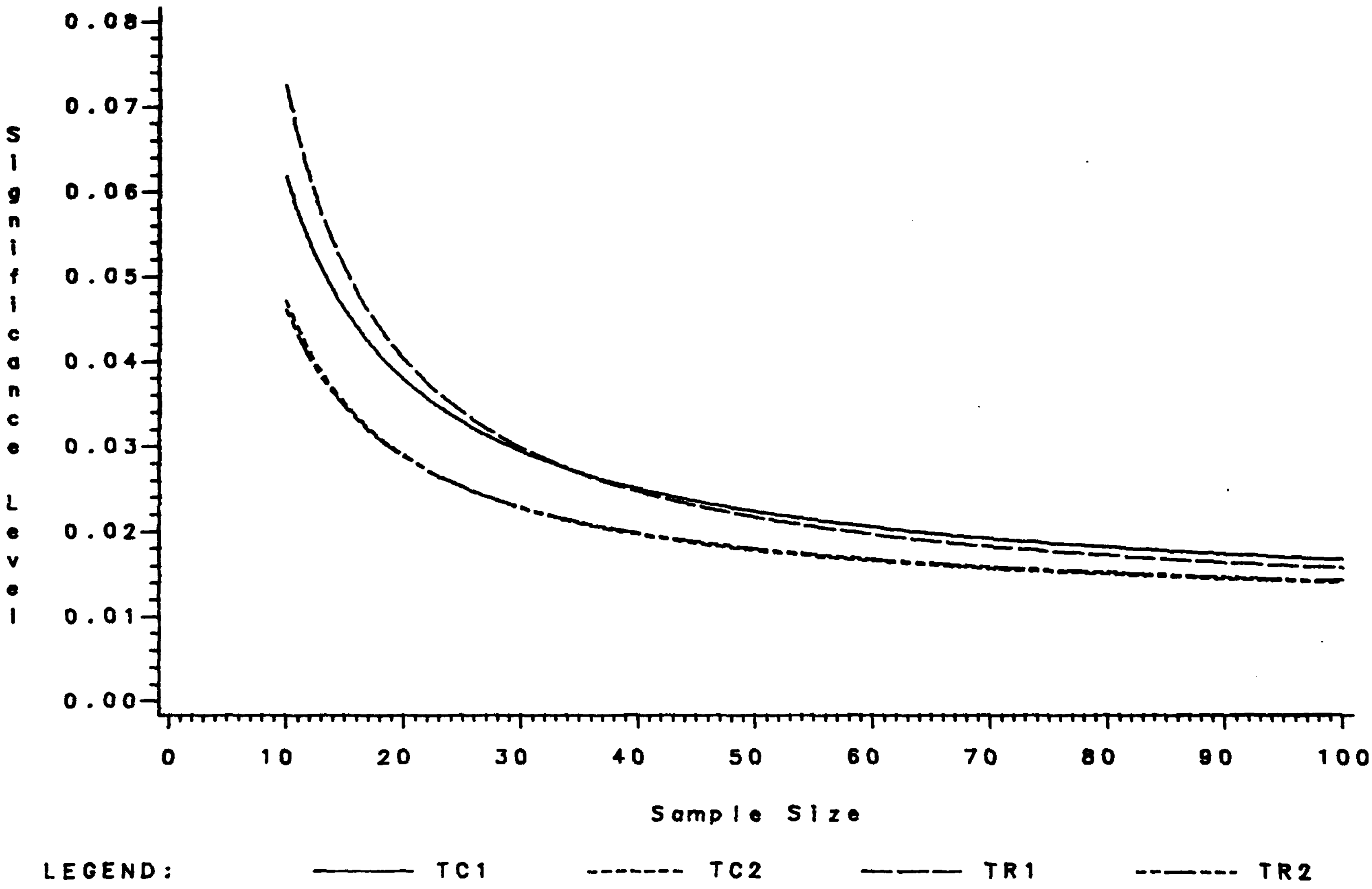


Figure 3: Empirical Significance Level as a Function of Sample Size for ρ=.5 and Nominal α=.01

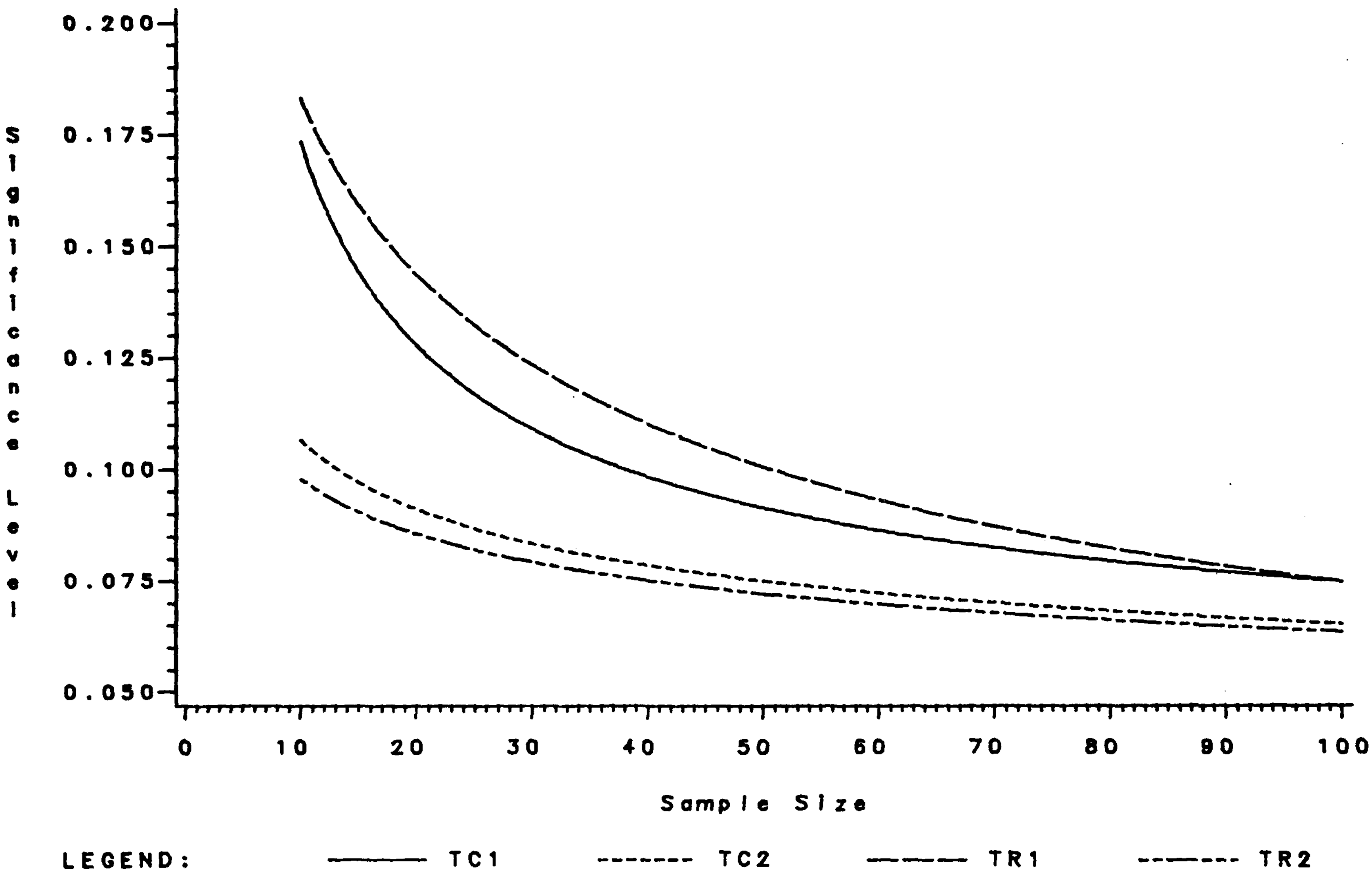


Figure 4: Empirical Significance Level as a Function of Sample Size for ρ=.7 and Nominal α=.05

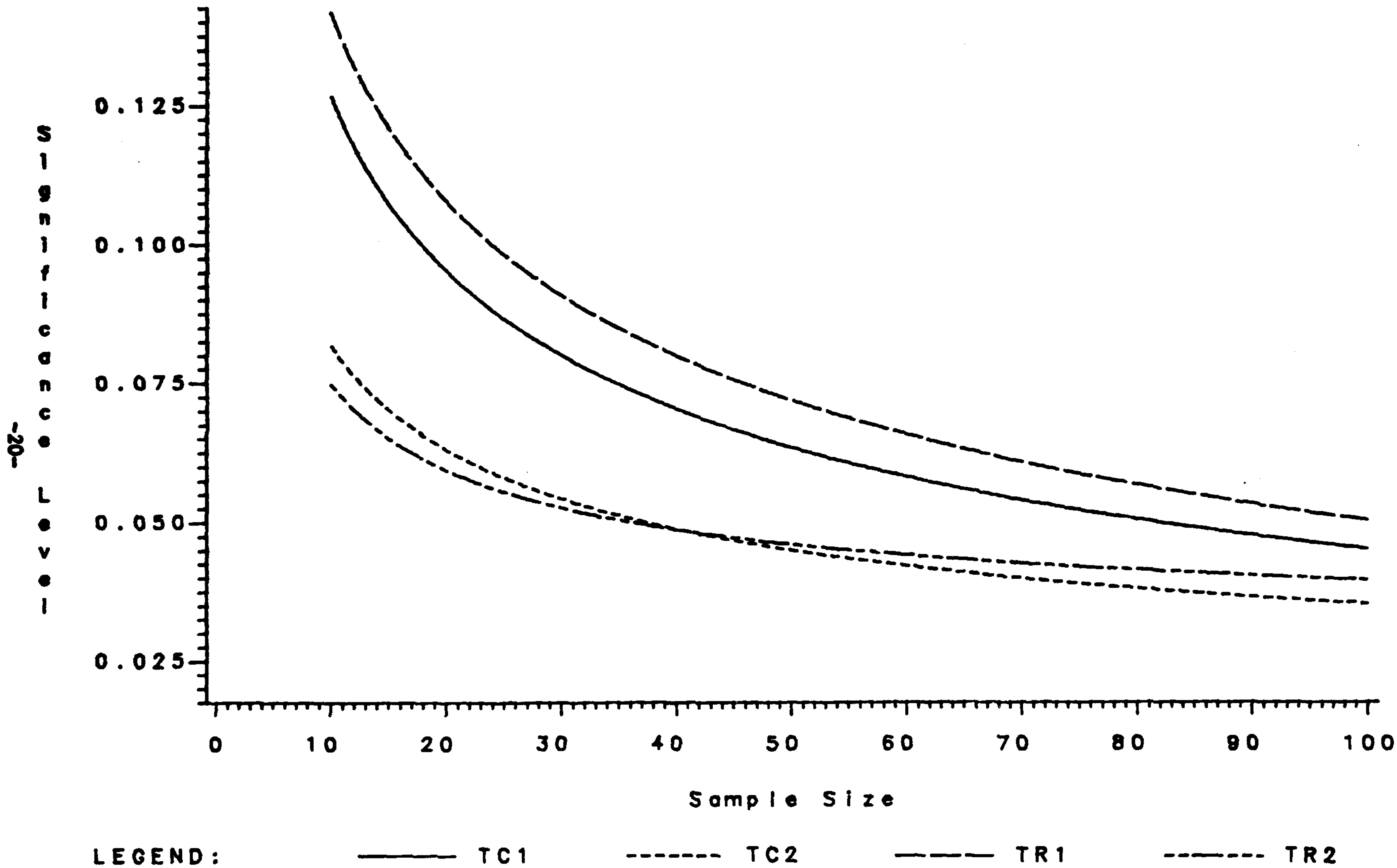


Figure 5: Empirical Significance Level as a Function of Sample Size for ρ=.7 and Nominal α=.025

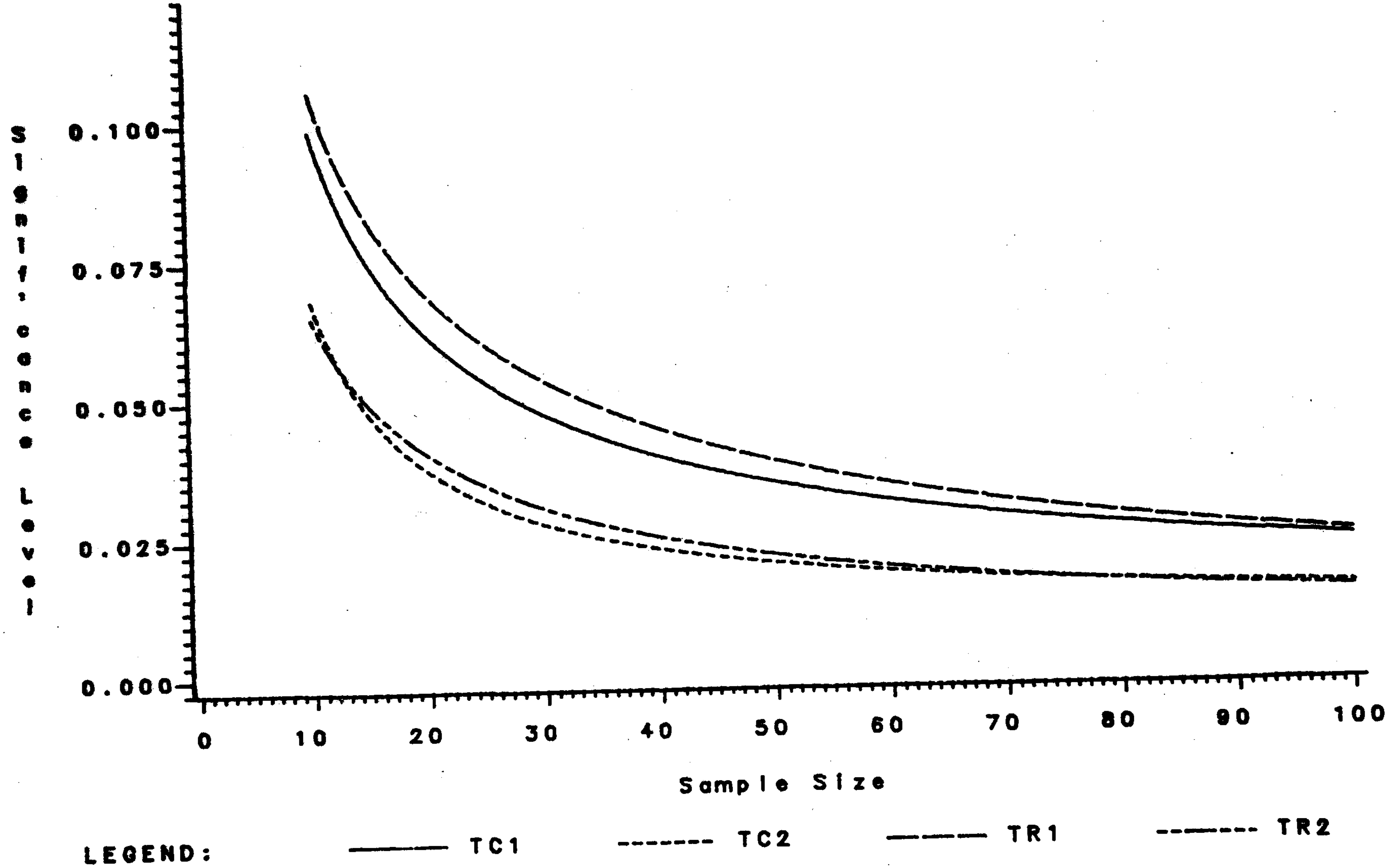


Figure 6: Empirical Significance Level as a Function of Sample Size for ρ=.7 and Nominal α=.01

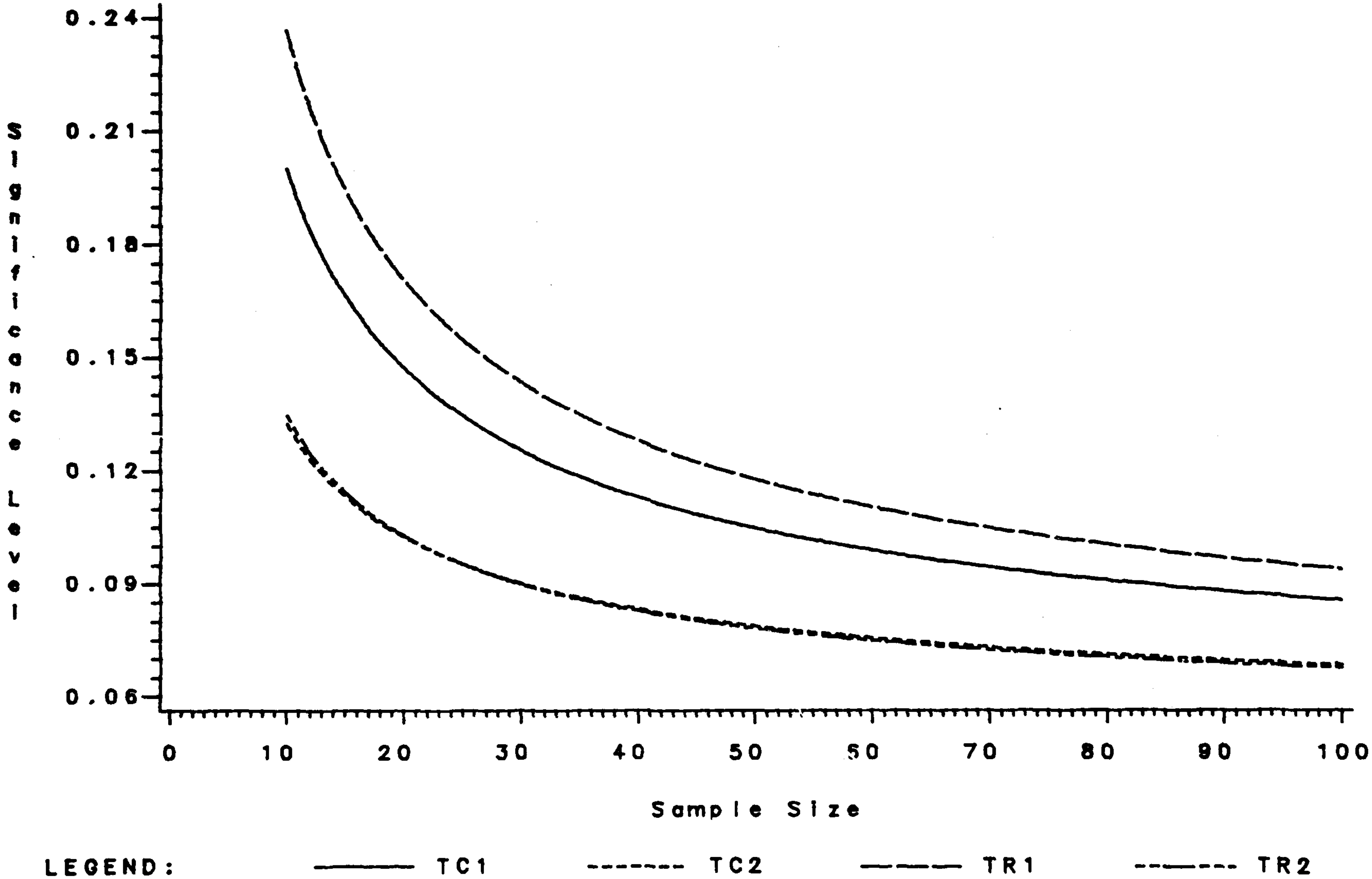


Figure 7: Empirical Significance Level as a Function of Sample Size for ρ=.8 and Nominal α=.05

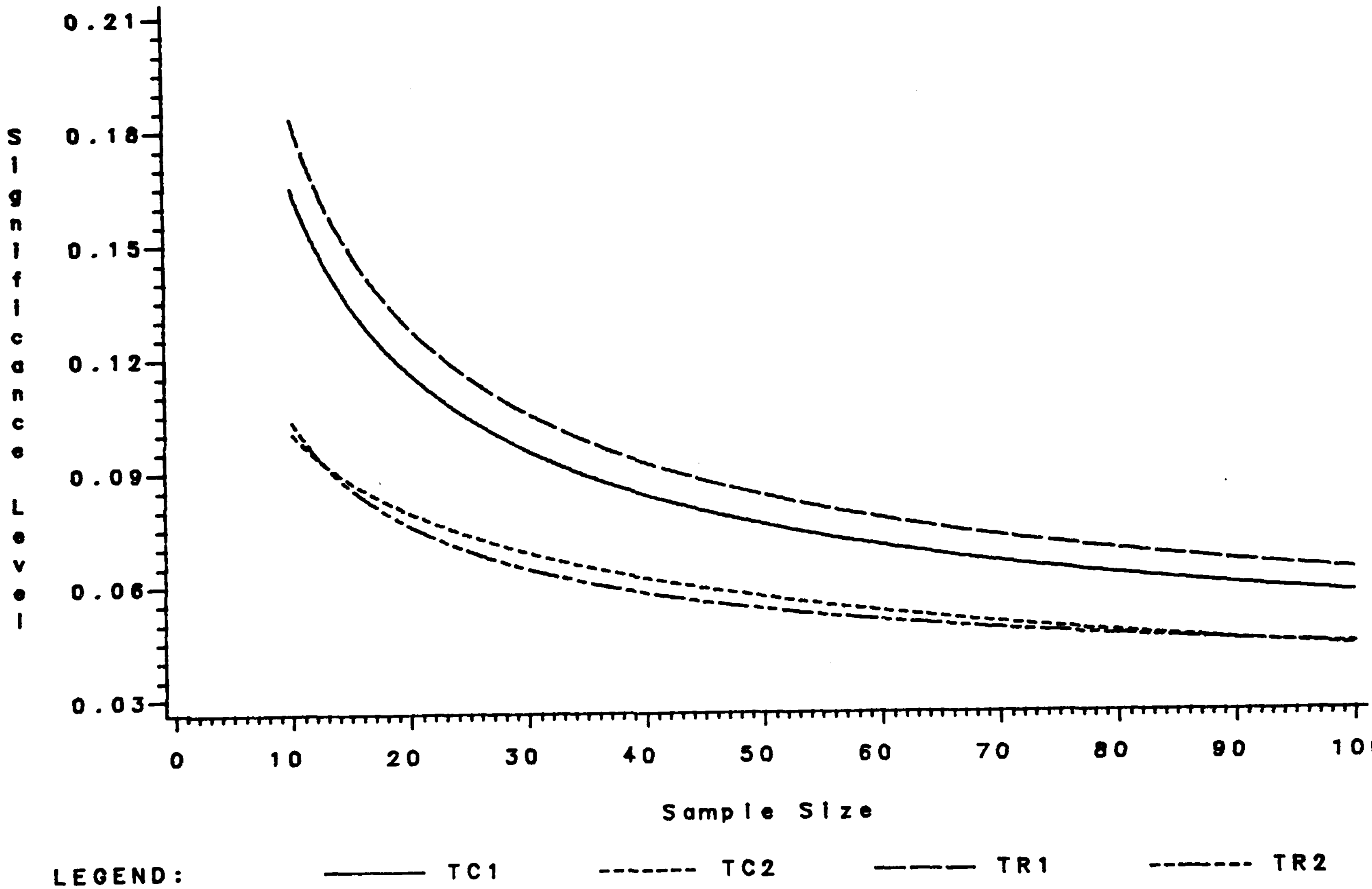


Figure 8: Empirical Significance Level as a Function of Sample Size for ρ=.8 and Nominal α=.025

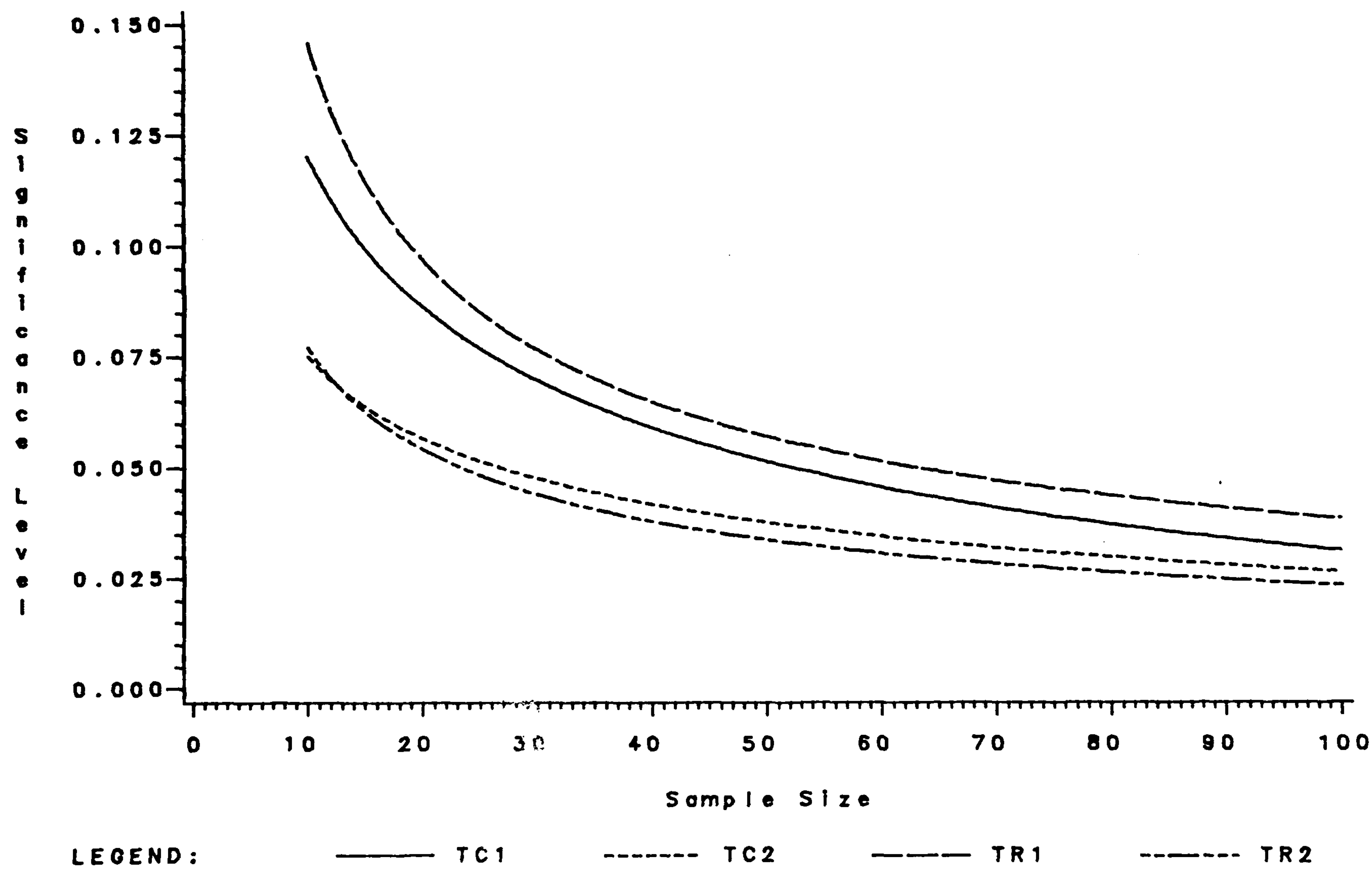


Figure 9: Empirical Significance Level as a Function of Sample Size for ρ=.8 and Nominal α=.01

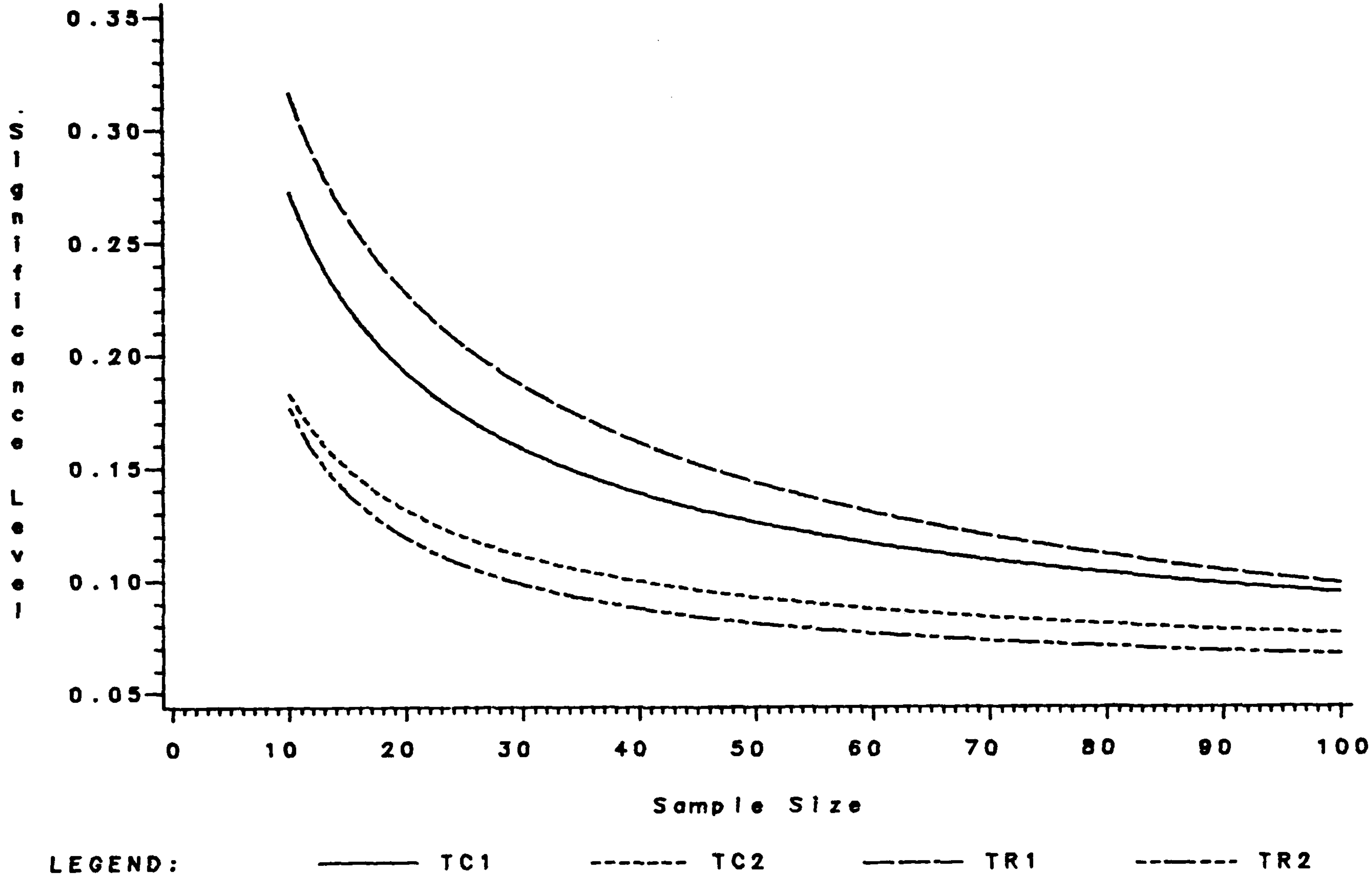


Figure 10: Empirical Significance Level as a Function of Sample Size for ρ=.85 and Nominal α=.05

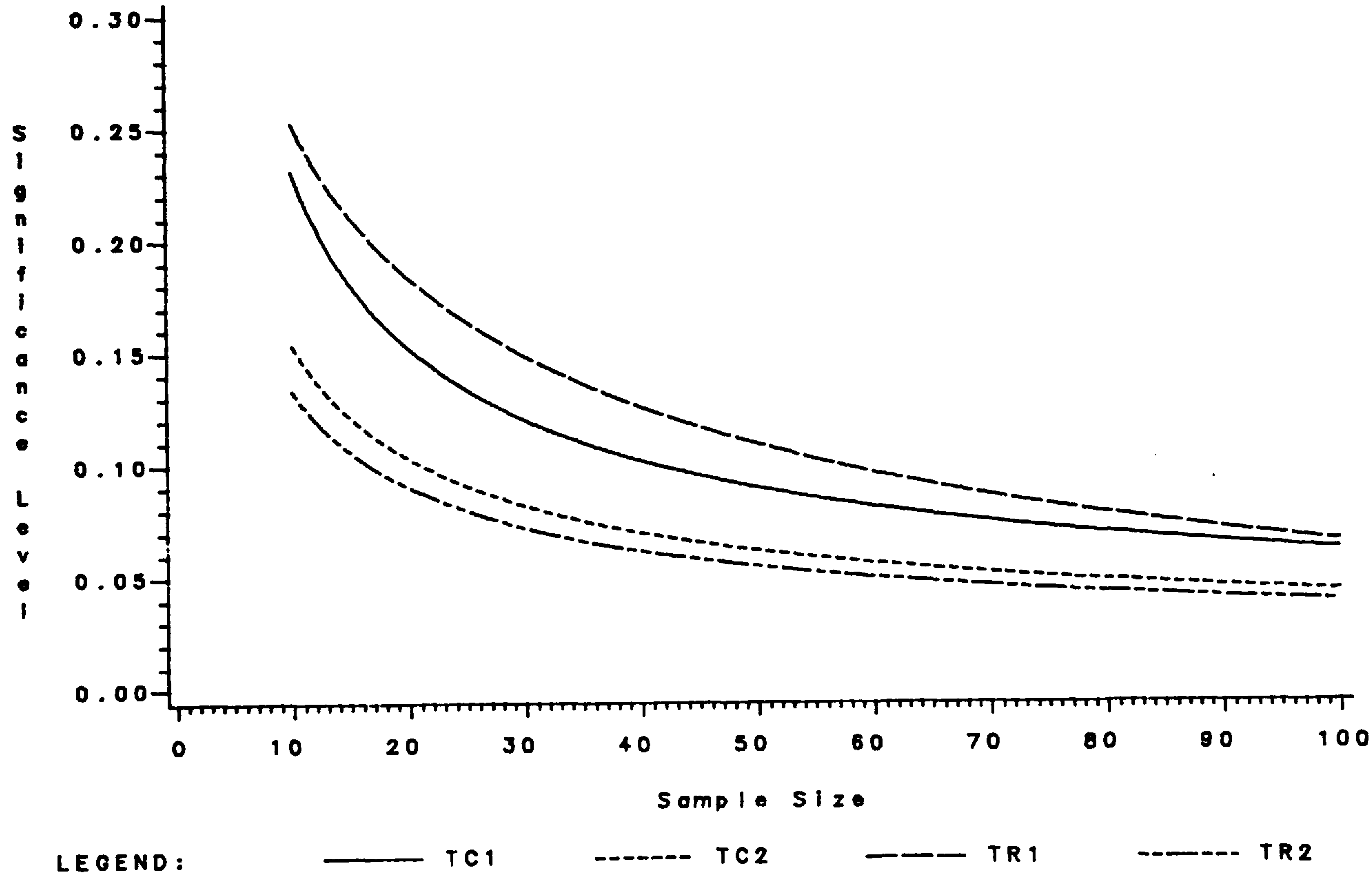


Figure 11: Empirical Significance Level as a Function of Sample Size for ρ=.85 and Nominal α=.025

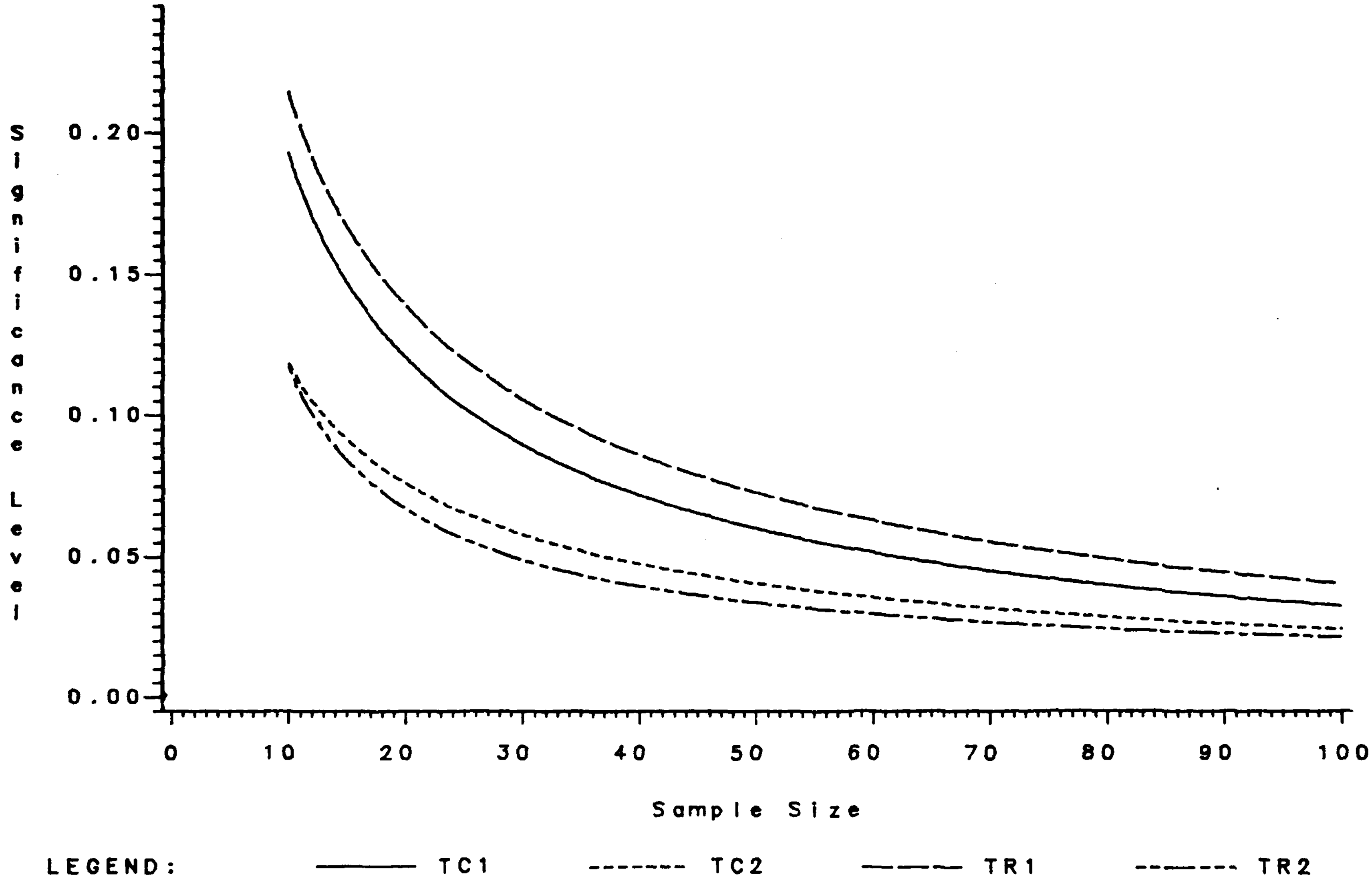


Figure 12: Empirical Significance Level as a Function of Sample Size for ρ=.85 and Nominal α=.01

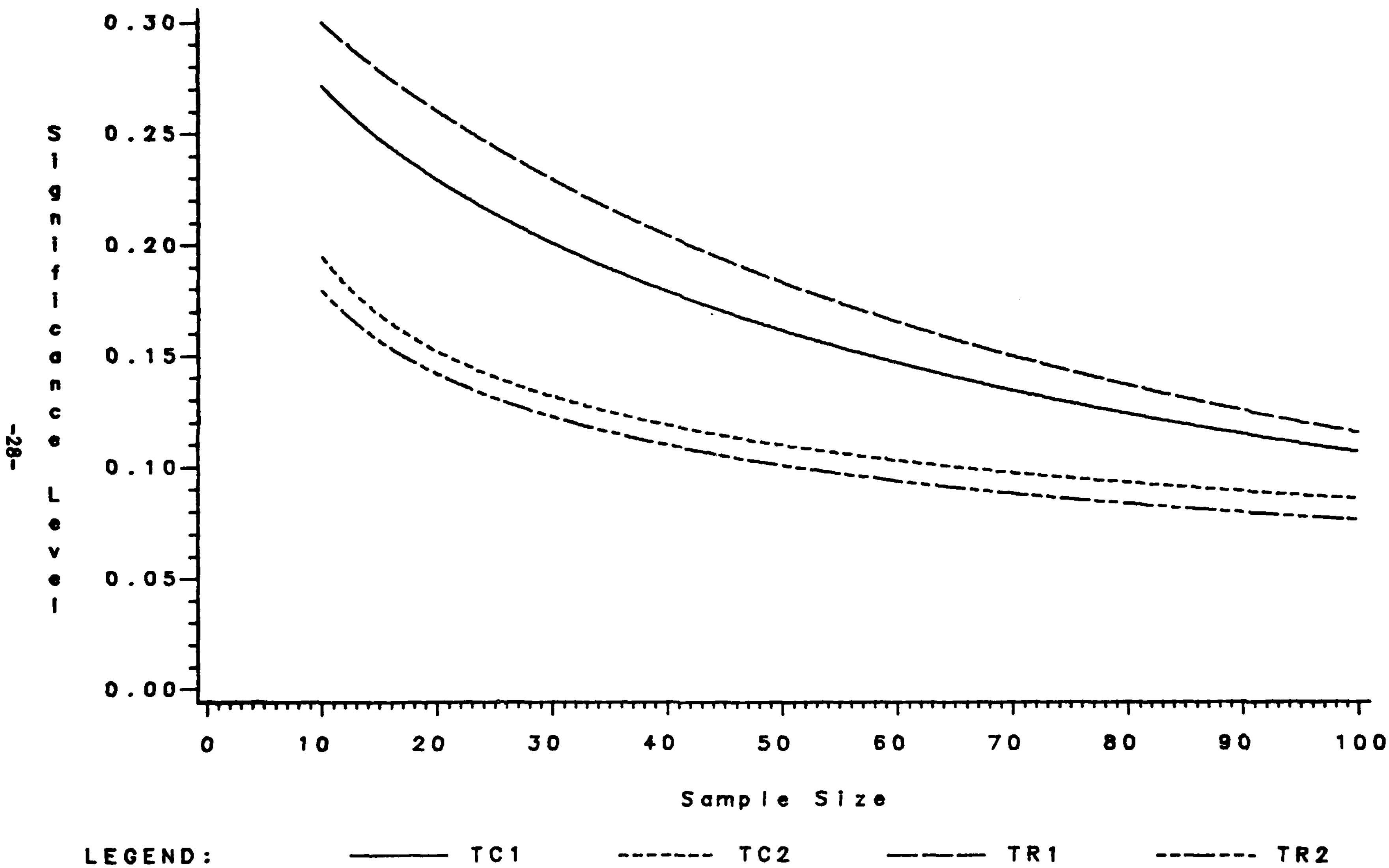


Figure 13: Empirical Significance Level as a Function of Sample Size for ρ=.9 and Nominal α=.05

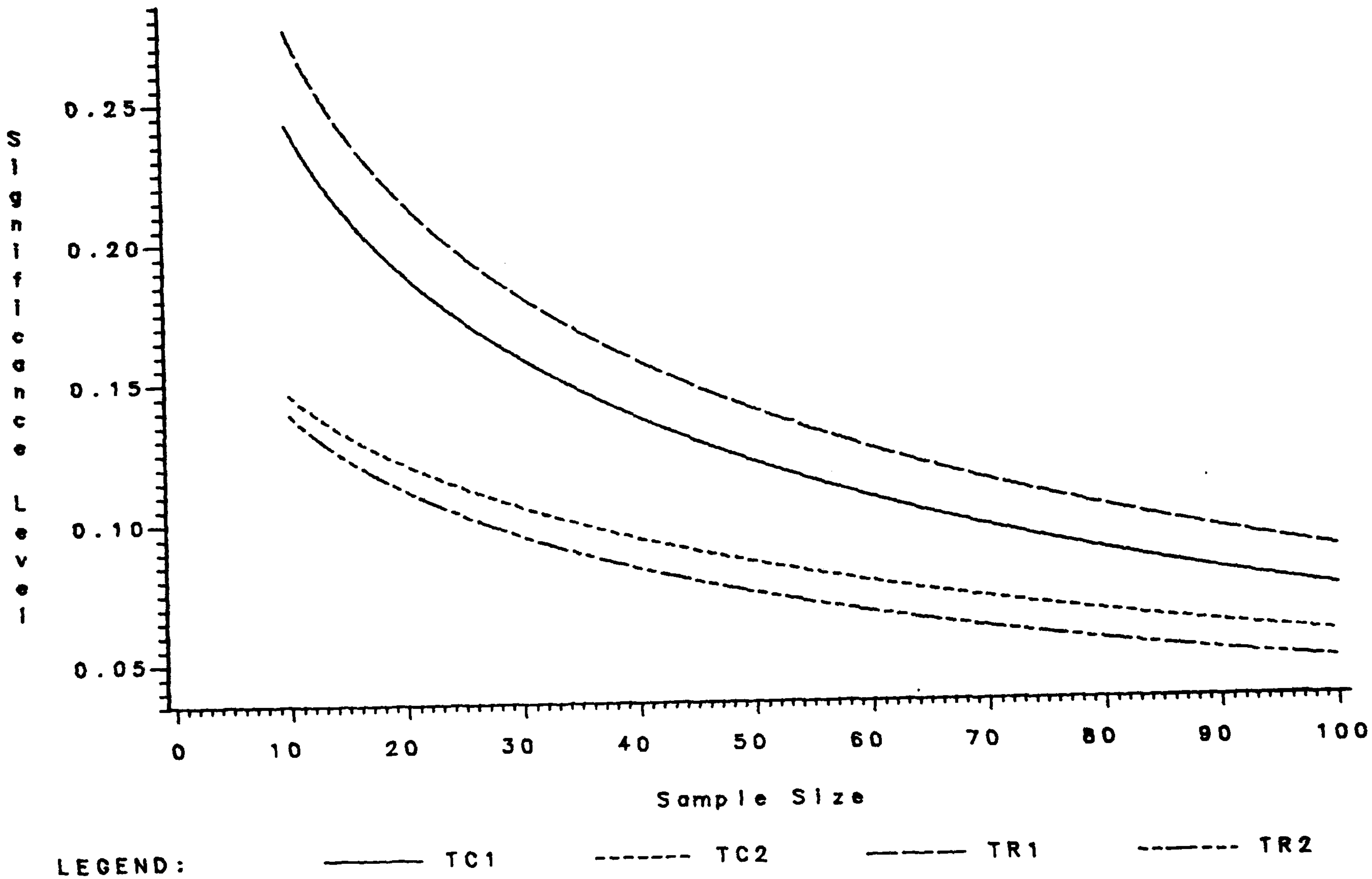


Figure 14: Empirical Significance Level as a Function of Sample Size for $\rho=.9$ and Nominal $\alpha=.025$

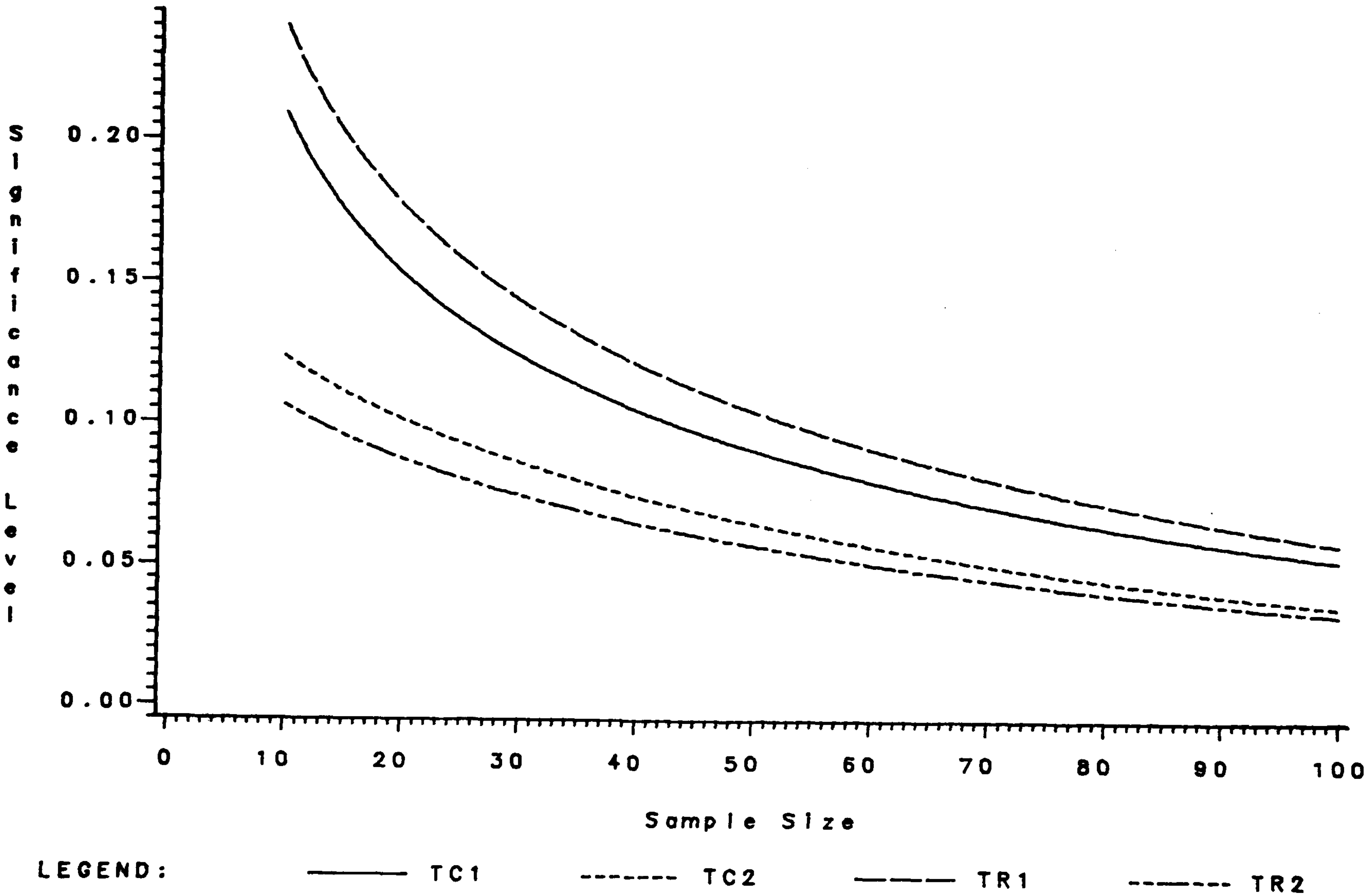


Figure 15: Empirical Significance Level as a Function of Sample Size for ρ=.9 and Nominal α=.01

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 112-14
2. GOVT ACCESSION NO.: AD-A129082
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): AN EMPIRICAL INVESTIGATION OF SEVERAL TESTS FOR THE MEAN OF A FIRST-ORDER AUTOREGRESSIVE PROCESS
5. TYPE OF REPORT & PERIOD COVERED: Technical Report
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Kevin C. Burns and Dennis E. Smith
8. CONTRACT OR GRANT NUMBER(s): N00014-79-C-0128
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Desmatics, Inc. P.O. Box 618 State College, PA 16801
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 207-037
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research Arlington, VA 22217
12. REPORT DATE: May 1983
13. NUMBER OF PAGES: 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Distribution of this report is unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
Autocorrelation
Single-Subject Experiments
AR(1) Model
20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

Four test statistics are considered for testing hypotheses about the mean of an AR(1) process. Simulated data are used to estimate the actual significance levels obtained when using $t_\alpha(n-1)$ as the critical value. Smoothing functions are fit to the empirical significance levels as functions of sample sizes. These functions are presented graphically and an example given as to how they might be used.